

# Exploring the Spectral Variability of the Earth as Measured by AVIRIS in 1999

Joseph W. Boardman\* and Robert O. Green\*\*

\*Analytical Imaging and Geophysics LLC, 4450 Arapahoe Ave., Suite 100, Boulder, CO

\*\*Jet Propulsion Laboratory, California Institute of Technology

## 1.0 INTRODUCTION

This short paper discusses our preliminary results from an initial exploration of the spectral variability of the Earth as exhibited in the 1999 AVIRIS archive. Unlike most AVIRIS investigations, where the investigator is trying to resolve detailed spatial and spectral questions about specific scenes, we used the 1999 AVIRIS archive as a whole to resolve larger scale, and more general, questions about AVIRIS itself, the data it produces and the spectral properties of the Earth. Our plan was to examine the second-order spectral statistics, means and covariances, of the 1999 AVIRIS archive in order develop quantitative data to address two types of questions: engineering questions and science questions. In this effort we only examined spectral statistics, no spatial analyses were done. A companion paper (Green and Boardman, 1999 this volume) examines the interplay of spatial and spectral properties of the data in terms of information content, spectral dimensionality and our engineering and science questions.

The engineering questions under investigation include:

- Considering the limited available resources, how is AVIRIS doing in terms of balancing its spectral spatial and radiometric performance levels?
- What can the AVIRIS data themselves tell us about how to improve the system?

The science questions we addressed, at least in part, include:

- What patterns exist in the AVIRIS data set as a whole?
- At current performance levels, what are the dimensionalities and spectral subspaces spanned by AVIRIS data?
- Can the world be split into classes of backgrounds?
- Which spectral regions and bands carry the information?
- Is AVIRIS truly “hyperspectral for all scenes?

Several papers by Price (Price 1994, Price 1997 and references therein) have questioned the need for high spectral resolution instruments, proposing that emphasis be placed in increasing spatial resolutions instead. The anecdotal experience of the authors, through many years of AVIRIS experience, has been to the contrary. One goal of the work described here was to examine this assertion by Price in detail, using high quality AVIRIS data to see if the hundreds of spectral channels truly provided useful information or are merely costly overkill.

The very definition of hyperspectral data implies “too many” bands. However this “hyper” state, where the instrument is more than capable, by virtue of its many bands, of capturing the dimensionality of the upwelling radiance field is a very desirable situation, not a bad one. The converse case is that of hypospectral data, where the instrument has fewer bands than the data have inherent spectral dimensions. In this sad situation our recorded data are then merely a “shadow” or subspace projection of the true radiance field and many possible exploitation methods and applications are precluded. It’s the difference between an overdetermined set of equations and an underdetermined one. Numerically we certainly desire a situation with “too many” bands in place of one with “too few”.

We are the first to admit that imaging spectrometers are not needed for all remote sensing applications. Certainly one does not need 224 spectral channels to perform coarse land cover classification tasks such as mapping gross urban, forest, water and vegetation classes. However, our experience has shown that when one is armed with hyperspectral data, versus hypospectral data, an entirely new realm of possible applications blossoms, out of the very overdetermined nature of the spectral data to which Price objects. The recent advances in spectral unmixing, nonlinear vegetation modeling and sub-pixel target detection algorithms bear witness to the outstanding results that are possible when hyperspectral data permit the merging of spectroscopic and imaging techniques. It is the “hyper” spectral leverage of “too many” bands that makes each of these developing applications possible.

This paper examines the spectral properties, in the form of means and covariances, of the 1999 AVIRIS archive, as individual scenes and as a global data set to examine these issues of dimensionality, overdeterminacy, AVIRIS performance and possible avenues for enhanced exploitation algorithms. In short we find that all 224 bands of AVIRIS provide useful information, that the best place to improve its performance is in the realm of radiometric precision and, assuming future noise level reductions, someday 224 bands may not be enough to guarantee hyperspectral data for complex Earth surface scenes.

## 2.0 PROCESSING METHODS

The processing consisted of two major portions. First, bulk processing was performed on the AVIRIS archive, collecting sums vectors and sums-of-squares matrices for the various flight lines of imaging spectrometry data. The sums vectors are essentially the totals of each spectral band. The sums-of-squares matrices are the totals of the various products formed when the 224 bands are multiplied by each other. The sums vectors have 224 elements and the sums-of-squares matrices are square with 224 elements on a side. Secondly, statistical processing was performed on the results of the bulk processing. The sums vectors and sums-of-squares matrices were converted to means vectors and covariance matrices. This conversion was performed on the scenes as individual units as well as on the pooled sums and sums-of-squares to derive global spectral statistics of the archive as a whole.

The bulk processing of the data was performed at JPL during off-hours at the AVIRIS data facility. A large portion of the 1999 AVIRIS archive was processed, including 510 flight strips. Both low and high altitude data were included. Overall, more than 1.14 billion spectra were processed and examined, amounting to more than 478 Gigabytes of raw data. The 510 scenes used were collected by AVIRIS between May 7 and November 14 of 1999 and range across the continent from the Bahamas to the Pacific Ocean. Many biomes were sampled including forest, oceans, deserts, cities and agricultural areas.

The conversion of the sums and sums-of-squares to means and covariances was performed for each of the individual scenes. In addition the 510 sets of single-scene bulk results were pooled to form a global sums and sums-of-squares matrix for full-archive analysis. These results were scaled to radiance units and then normalized by the 1999 noise equivalent delta radiance (nedL) spectrum. This scaling and normalization gave us mean spectra in units of noise standard deviations and covariance matrices in units of noise standard deviations squared. By scaling the results by the noise level we can quantitatively assess the signal levels and the data dimensionalities.

Finally each of the covariance matrices was subjected to an eigen analysis, deriving the associated eigenvalue vectors and eigenvector matrices. These results permit a direct examination of the various dimensionalities and subspaces spanned by the data sets as individual flight lines and as a global data set. To better model the noise in the data we fit linear functions to the noise-dominated tails of the eigenvalue spectra, rather than choosing a single threshold for signal versus noise cutoff. As a result we have more accurately modeled the signal dependent noise in the data and our data dimensionality estimates are made more conservative.

### 3.0 PROCESSING RESULTS

The 510 single scene results yielded 510 sets of mean spectra, covariance matrices, eigenvalue vectors and eigenvector matrices. Given the diversity of the many ground cover types, we found a wide range of spectral statistics. Figure 1 shows a typical single-scene sums vector and the associated sums-of-squares matrix. Figure 2 shows the results for the same single scene after conversion to the mean spectrum and covariance matrix form. Figure 3 shows two representations of the eigenvalues for this single-scene example. On the left the eigenvalues are shown on a linear scale. On the right, in a more informative plot, the same values are plotted on a log scale. Clearly the data dimensionality in this case exceeds forty, in terms of the number of spectral dimensions that have variance well above the noise level. The two plots show the trap that many, including Price, fall into when expressing dimensionality in terms of how many eigenvalues account for 95% or 99% of the data variance. Clearly just a handful of eigen dimensions are needed to explain 99% percent of the total variance. But this in no way means that the remaining 1-% is not useful information. On the contrary, we see that more than 35 subtle spectral dimensions are found in that last 1-% of the variance. It is truly the “baby in the bathwater” and should not be thrown out. It is this last 1-% of the spectral variance that is most important spectroscopically and that AVIRIS excels at capturing.

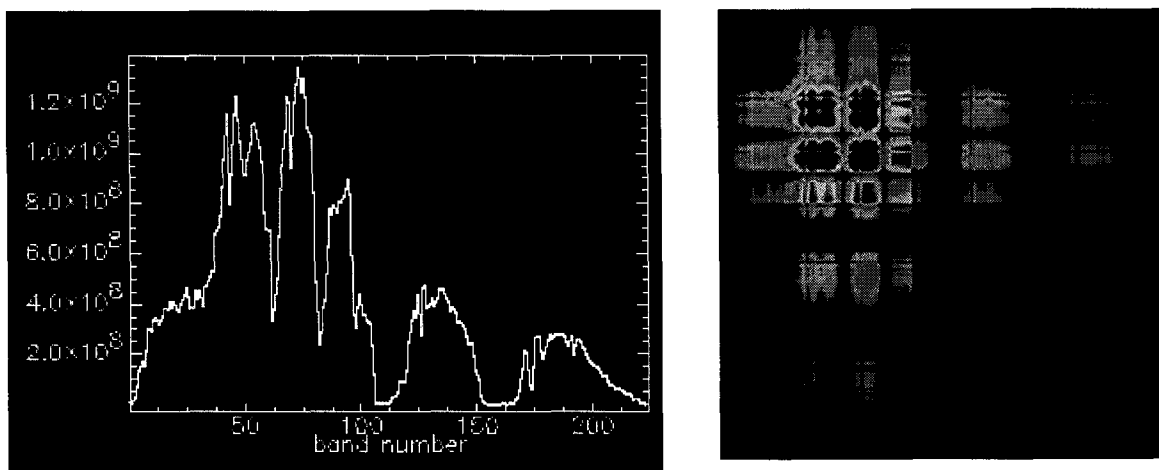


Figure 1. Example single-scene sums vector (left) and sums-of-squares matrix (right). Sums-of-squares matrix is color coded from high to low as red to black. Band 1 is in the upper left corner; band 224 is in the lower right.

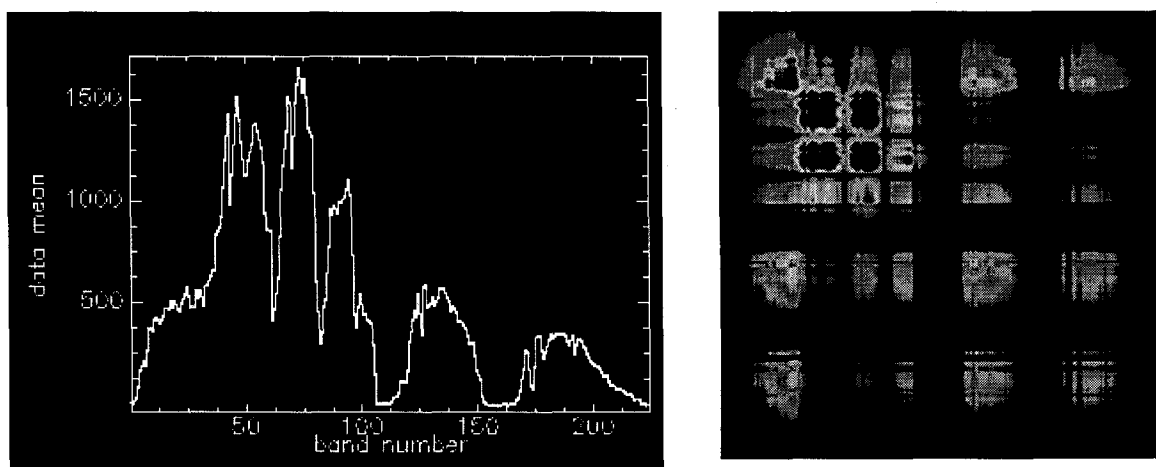


Figure 2. Example single-scene mean vector (left) and covariance matrix (right). Covariance matrix is color coded from high to low as red to black. . Band 1 is in the upper left corner; band 224 is in the lower right.

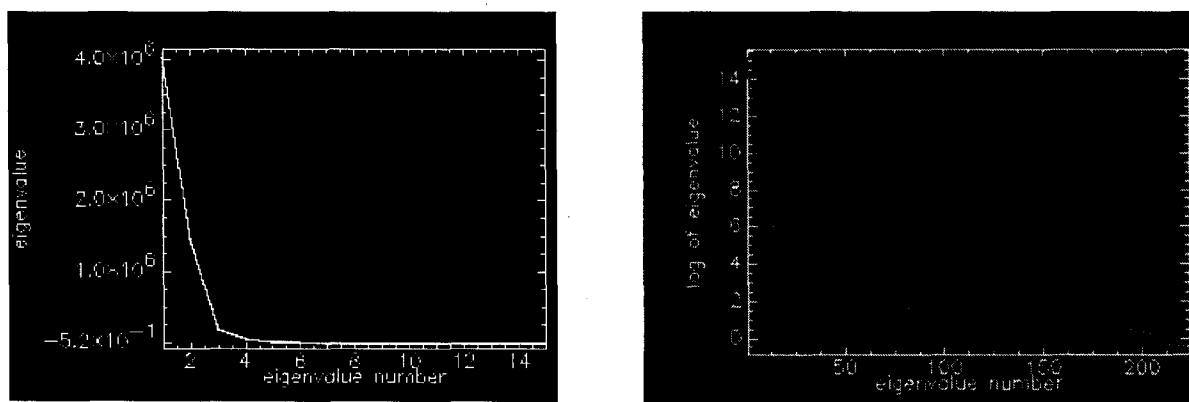


Figure 3. Example single-scene eigenvector plot, linear scale (left), log scale (right).

Figure 4 shows images of the associated eigenvectors for the same scene. The two image plots, one gray scale and the other color coded, depict the 224 224-element eigenvectors as rows of the image with eigenvector one at the top and eigenvector 224 at the bottom. The break in character in the eigenvectors, at approximately eigenvector 40 or 50, is a clear depiction of the break between signal space and noise space. The remaining eigenvectors show the pattern of the SVD algorithm simply filling in the null space with a progression of orthogonal vectors.

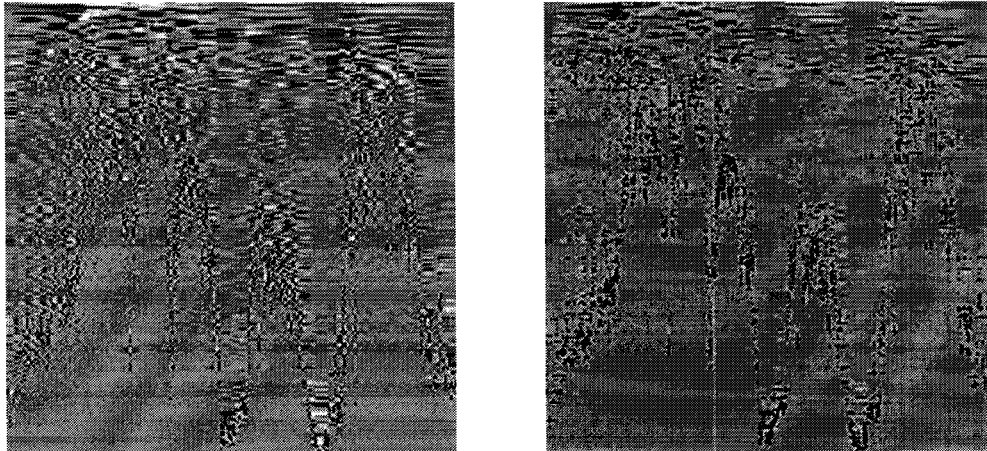


Figure 4. Gray scale (left) and pseudocolor (right) depictions of the eigenvectors for a single-scene example. Eigenvectors are rows in these images and are ordered top to bottom.

Measuring dimensionality in multivariate data sets is a slippery slope and an approximation at best. We decided to err on the side of conservatism and tend towards underestimation of dimensionality. To this end we used a conservative noise model and chose to cut off the dimensionality estimates at different multiples of the noise level. Figure 5 shows the results for dimensionality determination for the 510 single scenes. The pseudocolor display shows a column for each scene, color coded with a stacked indication of how many eigenvalues exceed various threshold multiples of the noise level. Variances 1000, 100, 10, 5 and 3 times the noise level are shown in red, yellow, green blue and magenta respectively. The white horizontal bars are for reference and dimensionality determination. They are spaced 20 dimensions apart, starting at zero on the bottom, with white bars at dimensionalities of 20, 40, 60, 80 and 100 as you go up. Clearly, by anyone's measure, AVIRIS is measuring some tremendously high-dimension data, with many scenes having more than 60 or even 80 dimensions with variance more than three times the ambient noise level. Nevertheless not a single scene has 224 eigenvalues that exceed the noise cutoff that would render the data hypospectral. Clearly AVIRIS has more "room" to improve its radiometric precision before it needs to add bands. However the current number of bands is not more than twice the number required by the current nedL levels to maintain hyperspectral leverage. Figure 6 shows the histogram of the distribution of scene dimensionalities, where we used a threshold of three times the noise level. Obviously AVIRIS measured many scenes in 1999 that exceed 50 dimensions of unique spectra information, even with this conservative metric.

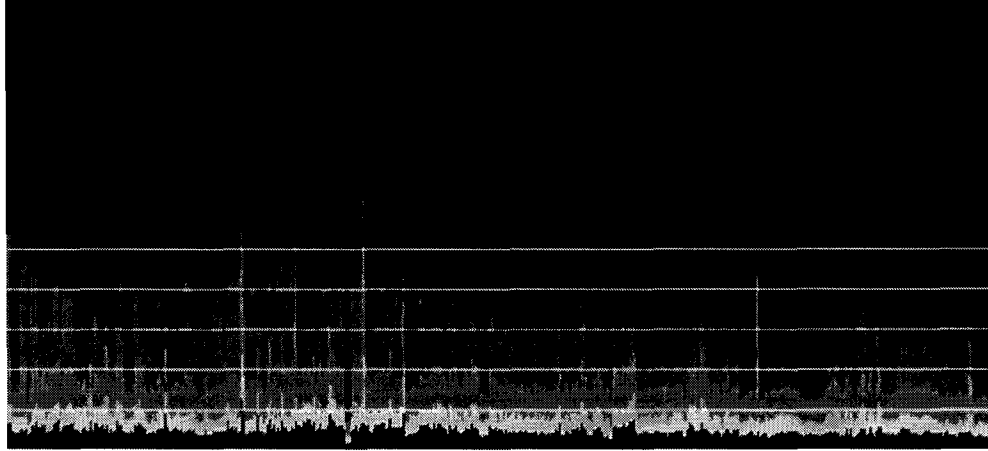


Figure 5. Spectral dimensionality determination for the 510 scenes processed. See text for discussion.

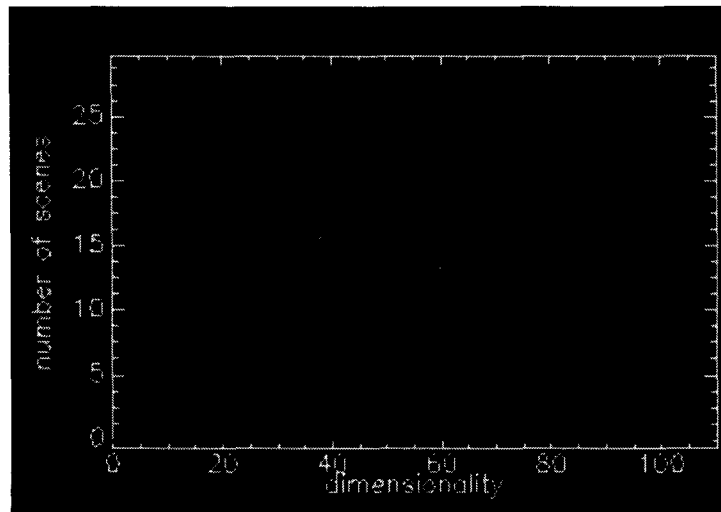


Figure 6. Histogram of the single-scene dimensionalities for a conservative 3-time-noise level threshold.

The covariance matrices, along with the mean spectra, define the span of the spectral variation in the 510 scenes processed. Treating these data sets as metrics we can measure the distance among them to classify the scenes into groups that share the same spectral statistics and data span. Figure 7 shows a matrix that measures the spectral “distances” among the various covariance matrices. The “tartan-plaid” appearance relates to the several gross classes of scenes including forest, open water, deserts etc. Each class of spectral properties has its own characteristic covariance matrix and thus its own characteristic distances to the other covariance matrices. We are currently investigating the use of these metrics for meta-clustering of scenes and development of robust spectral background classes and their relevant spectra statistics and characteristic subspaces.

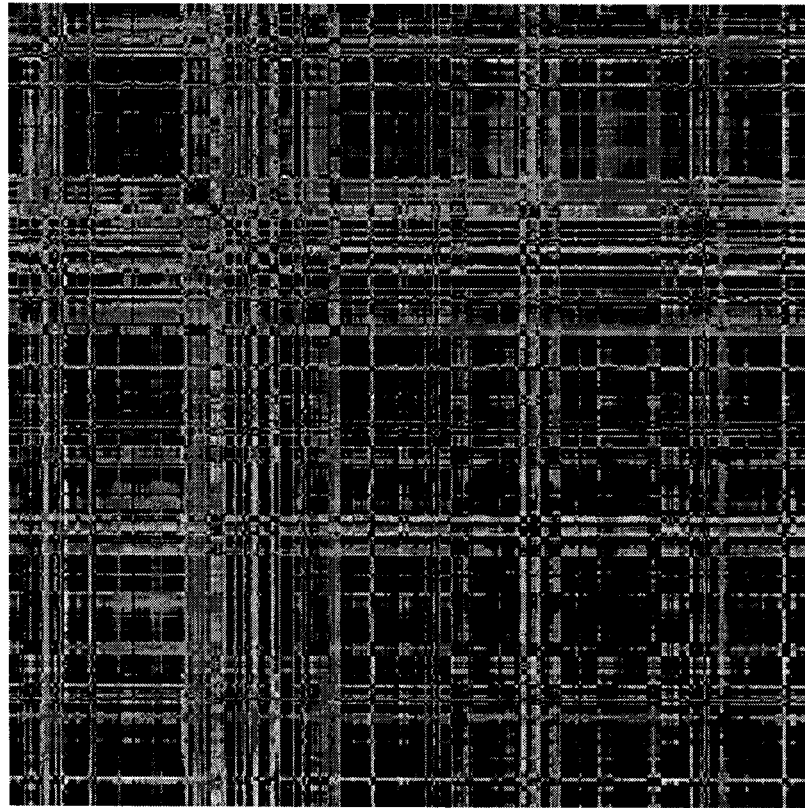
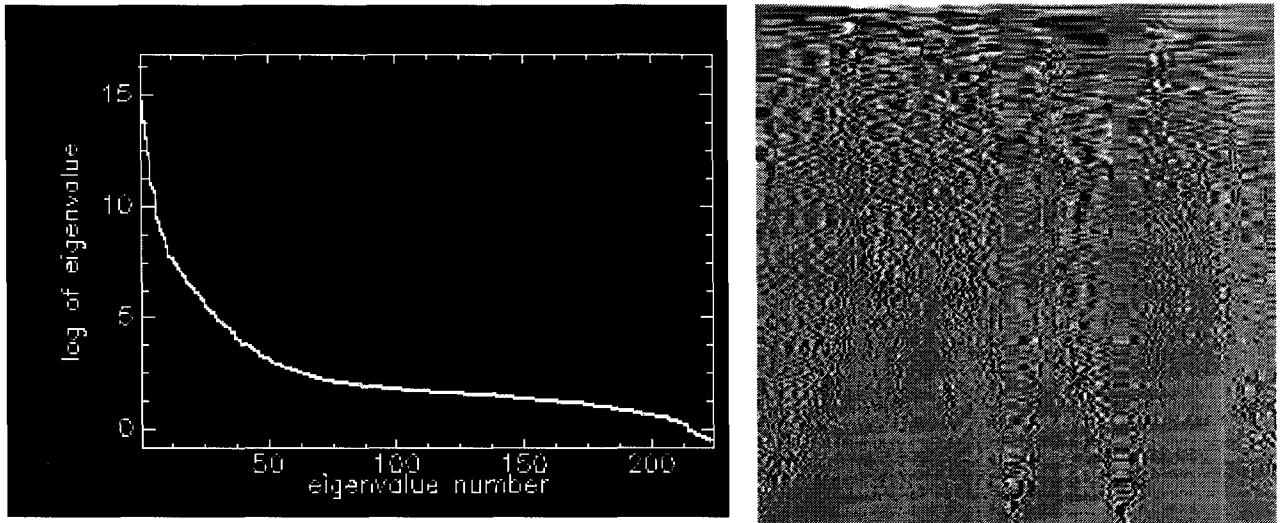


Figure 7. Inter-scene covariance separation metric matrix, showing the diversity of the spectral subspaces in the 1999 archive and the gross clustering of robust background classes. Larger distances to smaller distances are color coded as red to green to blue to black.

Combining the single-scene raw results into a global data permits the examination of the overall spectral space image by AVIRIS in 1999 as a whole, across all 510 scenes and across the entire continent. The global scene begins to approximate what an on-orbit sensor would be faced with and will be useful for system design trade studies. Figure 8a shows the log plot of the global eigenvalues. The dimensionality exceeds 60 or 70. Since this estimate is biased towards common materials and rare endmembers have a hard time driving such a large pooled covariance, we expect the actual span of the global subspace imaged by AVIRIS is considerable higher, probably in triple digits. Figure 8b shows the associated eigenvector matrix image. Notably there is no strong break between signal and noise spaces. This pooled covariance matrix results in an eigenvector set that exhibits signal characteristics well out past 60 dimensions. Figure 9 shows individual spectral plots of the first nine eigenvectors of the pooled covariance matrix. Notable in these plots is the prevalence of many narrow spectral features, beyond those associated with the atmosphere and sun, due to subtle spectral signals in the surface mineral, vegetation and man-made cover types. Clearly all 224 bands of AVIRIS contribute information and no part of the spectrum is insignificant.



Figures 8a and 8b. Global eigenvalue plot, log scale and the associated global eigenvector matrix image as a gray scale display.

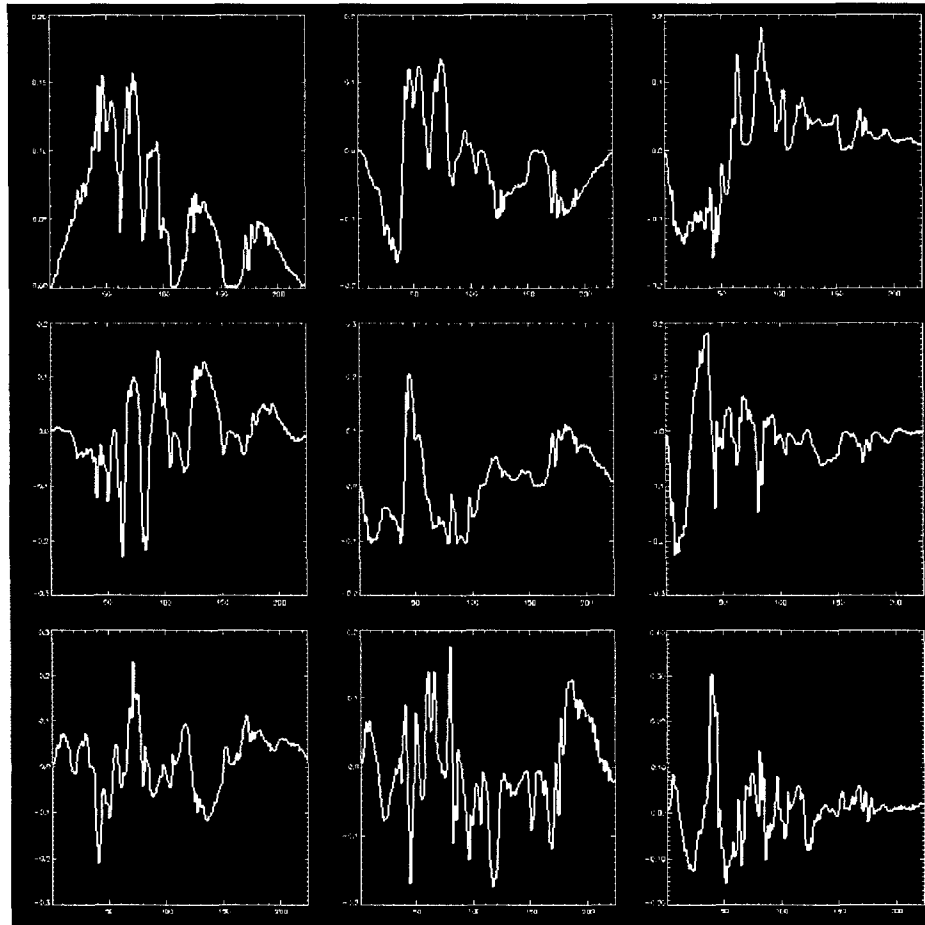
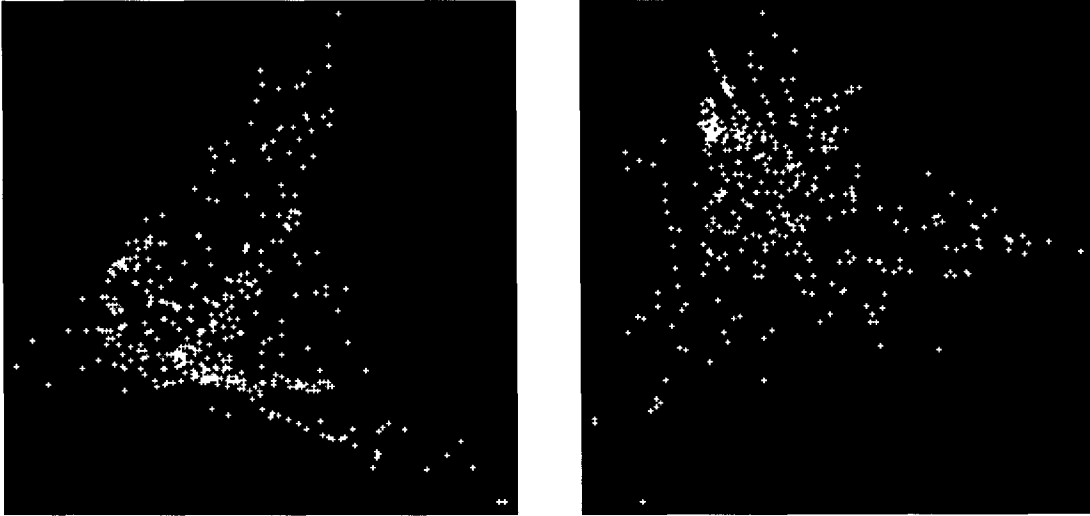


Figure 9. First nine global eigenvector plots showing the narrow spectral features spanned by the data subspace and the spread of information across the spectral range.



Projecting each of the 510 single-scenes onto the global eigenvectors allows us to examine meta-clusters of scenes and the global distribution of reflectances. Figures 10a and 10b show the projection of the single scenes onto the first three eigenvectors of the global covariance matrix. Interactive examination of this type of data visualization has revealed many fascinating patterns of the distribution of reflectances across the continent. We are currently examining the results from this unique visualization of over one billion spectra to extract AVIRIS performance and surface reflectance results.



Figures 10a and 10b. Projections of the 510 scene means onto the first and second (left) and first and third (right) eigenvectors of the global covariance matrix. Global patterns of vegetation, water and mineral materials are exposed in this unique visualization of over one billion spectra.

#### 4.0 INSTRUMENT AND ENGINEERING CONCLUSIONS

From our preliminary investigations of the statistics of the 1999 AVIRIS archive we can draw several important conclusions about the system and its performance.

- AVIRIS continues to be hyperspectral “despite” its greatly enhanced nedL performance, however in some cases the single-scene dimensionality is more than half the number of bands.
- The full range of spectral bands contributes to the natural spectral span of the data, with many eigenvector features having full-width-half-max features of less than 50 nm in width.
- NedL continues to be the limiting parameter in terms of data dimensionality and thus information content. We are still not to a natural floor in signal and our information floor is a noise floor.
- Spectral “room” remains for considerable increases in nedL performance where possible.

- Experimental studies are indicated to examine how increased signal-to-noise-ratios can be achieved through increased integration times (flying slower) or by trading spatial for radiometric resolutions (flying lower and binning pixels, see Green and Boardman, 1999 this volume).

## 5.0 SCIENCE AND EXPLOITATION CONCLUSIONS

The science and exploitation conclusions are manifold and we are reporting only preliminary results here.

- Observable radiance fields regularly exceed dimensionality 40 and some even exceed 100 dimensions, given AVIRIS' current nedL performance. This is proof of the extreme information content in our data and a real challenge for the development of robust exploitation methods.
- Global and single-scene covariance statistics define spectral subspaces occupied by Earth radiances and the AVIRIS archives are a unique resource for this important study and modeling effort.
- Clever uses of a priori knowledge will permit real-time, on-board target detection and feature mapping and possibly obviate standard atmospheric data reduction and off-line processing systems currently in place.
- We've only scratched the surface of possible scientific and commercial applications that are possible with AVIRIS-class instruments.

## 6.0 REFERENCES

Price, John C., 1994, Band selection procedure for multispectral scanners, *Applied Optics*, vol. 33, no. 15, pp. 3281-3288.

Price, John C., 1997, Spectral band selection for visible-near infrared remote sensing: spectral-spatial resolution tradeoffs, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, no. 5, pp. 1277-1285.

## 7.0 ACKNOWLEDGEMENTS

A large portion of this research was carried out at the Jet Propulsion Laboratory, California Institute of technology, under contract with the National Aeronautics and Space Administration. The authors would like to express their appreciation for the efforts of the AVIRIS team at the Jet Propulsion Laboratory. Additionally, Analytical Imaging and Geophysics provided funding and support from internal sources for this effort.